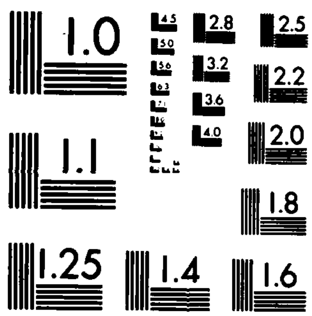
AD-A129 746 A LIKELIHOOD RATIO TEST REGARDING TWO NESTED BUT OBLIQUE ORDER RESTRICTED..(U) IOWA UNIV IOWA CITY DEPT OF STATISTICS AND ACTURIAL SCIENCE.. G WARRACK ET AL. 1/1

UNCLASSIFIED NOV 82 TR-90 N00014-80-C-0321 F/G 12/1 NL

END
DATE
FILMED
83
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0

(12)

AD A123746

A LIKELIHOOD RATIO TEST REGARDING
TWO NESTED BUT OBLIQUE ORDER
RESTRICTED HYPOTHESES

Giles Warrack and Tim Robertson

Department of Statistics
and Actuarial Science
The University of Iowa
Iowa City, Iowa 52242

Technical Report #90

American Mathematical Society 1979 subject classification

Primary 62F03 Secondary 62E15

Key words and phrases: Order restricted inference, likelihood ratio tests, unipolar affective disorder, chi-bar-square distribution, isotonic regression.

* This research was sponsored by the Office of Naval Research under Contract N00014-80-C-0321. It is taken from the first author's doctoral dissertation at The University of Iowa.

A LIKELIHOOD RATIO TEST REGARDING TWO NESTED BUT OBLIQUE ORDER RESTRICTED HYPOTHESES

Giles Warrack and Tim Robertson

SUMMARY

In an article in the Journal of Psychiatric Research, [Cadoret, Woolson and Winokur, 1977] consider two theories regarding the genetic makeup of patients suffering from unipolar affective disorder. These two theories imply nested but oblique order restrictions on the parameters of a statistical model. A likelihood ratio test for these two restrictions is studied.

1. Introduction.

In this paper we consider a hypothesis test in which both the null and the alternative hypotheses impose an order restriction upon a set of parameters. Specifically, we deal with a situation in which we take independent samples from k populations, where each population is distributed according to a particular member of an exponential family (we also consider sampling from a multinomial population). If we denote by θ_i the parameter of interest from the i^{th} population, we consider hypotheses H_0 and H_A, where

(1.1) H_0: $\theta_1 = \theta_2 = \cdots + \theta_{k_1} \geq \theta_{k_1+1} = \cdots = \theta_{k_2} \geq \cdots \geq \theta_{k_{\ell-1}+1} = \cdots = \theta_k$

(1.2) H_A: $\theta_1 \geq \theta_2 \cdots \geq \theta_{k_1} > \theta_{k_1+1} \geq \cdots \geq \cdots \geq \cdots \geq \theta_k$.

We note that H_A contains H_0, and we wish test H_0 against H_A-H_0 (H_A but not H_0).

This investigation was stimulated partly by a problem encountered in psychiatric research. [Winokur et al., 1971] studied data on psychiatric illnesses afflicting family members of patients suffering from unipolar affective disorder (u.a.d.). This data strongly suggested a higher rate of certain kinds of illness in family members of patients who were affected by u.a.d. early in life than the rate for family members of patients for which the onset of the disease occurred later in life. This research suggested genetic differences between these patients which could be tested using data on psychiatric illnesses suffered by family members. In Cadoret et al. two different explanations (theories) are examined which might explain this phenomenon. The first postulates a qualitative difference whereby there exist two qualitatively different genes or groups of genes which

explain the differences between the early onset probands and the late onset probands. The second postulates differences caused by a single group of genes with differences between the groups due to the number of genes (the quantitative theory).

Cadoret et al. (1977) examined data on 6 groups of patients suffering from u.a.d. The groupings are based upon the age at which the patient is first affected by the condition. Conclusions regarding the genetic makeup of these patients were based upon inferences about the parameters p_i; $i = 1,2,\ldots,6$ where in one case p_i represents the proportion of alcoholic fathers of patients in the i^{th} age group and in another case it represents the proportion of depressive parents of patients in the i^{th} group. In terms of these parameters the researchers quantify the two theories as follows:

$$H_0: \; p_1 = p_2 = p_3 \geq p_4 = p_5 = p_6$$

$$H_A: \; p_1 \geq p_2 \geq p_3 \geq p_4 \geq p_5 \geq p_6$$

where H_0 represents the qualitative theory and H_A represents the quantitative theory. The problem is to use the data to decide between H_0 and $H_A - H_0$; thus this problem falls within the general framework we are considering.

The hypothesis test represented by (1.1) and (1.2) is a special case of a large class of problems in statistics where both the null hypothesis and the alternative impose an order restriction on the parameter set. Additional research on order restricted hypothesis tests where the null hypothesis is composite can be found in [Robertson and Wright, 1981] and [Dykstra and Robertson, 1982,1983]. An interesting feature of

the test regarding (1.1) and (1.2) is the null hypothesis behavior of the distribution of the likelihood ratio statistic which is not what one might expect in light of previous research in the area of testing statistical hypotheses under order restrictions. In virtually all of these previously investigated problems either the null hypothesis is a similar region for the likelihood ratio test (cf. Barlow et al. (1972)) or homogeneity is the least favorable configuration within the null hypothesis for the distribution of the likelihood ratio test (cf. [Robertson and Wegman, 1978, Robertson, 1975 and Robertson, 1978]). By homogeneity being "least favorable" we mean that for all possible values of the parameter set, within the null hypothesis region, the likelihood ratio statistic achieves its supremum when all the θ_i are equal. We, however, obtain a rather surprisingly different result which is described below.

Suppose that the θ_i are the means of normal populations with known variances, and suppose that based upon independent samples, we wish to test H_0 against $H_A - H_0$ where, say

$$H_0: \ \theta_1 = \theta_2 = \cdots = \theta_{k_1} \geq \theta_{k_1+1} = \cdots = \theta_k$$

$$H_A: \ \theta_1 \geq \theta_2 \geq \cdots \geq \theta_{k_1} \geq \theta_{k_1+1} \geq \cdots \geq \theta_k .$$

Then we are faced with the problem that the likelihood-ratio based statistic, L, is not similar over the null hypothesis region. This time the $\theta \in H_0$ giving $\sup_{\theta \in H_0} P_\theta[L > t]$ is that vector such that the difference between the first k_1 components and the second $k-k_1$ components approaches infinity. Furthermore, this result enables us to show that

(1.3) $\quad \sup_{\theta \in H_0} P_\theta[L > t] = P[L' > t]$

where L' is the likelihood ratio based statistic for testing H_0' against $H_A' - H_0'$ where

(1.4) H_0': $\theta_1 = \dots = \theta_{k_1}$; $\theta_{k_1+1} = \dots = \theta_k$

(1.5) H_A': $\theta_1 \geq \dots \geq \theta_{k_1}$; $\theta_{k_1+1} \geq \dots \geq \theta_k$

The statistic L' is shown to be distributed as the sum of two independent random variables each of which is distributed according to Bartholomew's chi-bar-squared distribution. A similar result is obtained in dealing with hypotheses of the form (1.1) and (1.2), in which we test for a descending trend between ℓ sets of parameters, with equality within the sets, or "blocks", against the hypothesis of a descending trend both within and between blocks.

In Section 2 we consider tests for the normal distribution, and in Section 3 we use these results to derive asymptotic tests for the exponential class of distributions and the multinomial distribution. In Section 4 we illustrate the use of the techniques developed here by analyzing one of the data sets in [Cadoret et al., 1977].

2. Tests for the normal distribution.

We begin by considering the case in which we sample from k independent normal populations with known variances. We wish to test H_0 against $H_A - H_0$ where

(2.1) H_0: $\mu_1 = \mu_2 = \dots = \mu_{k_1} \geq \mu_{k_1+1} = \dots = \mu_k$

(2.2) H_A: $\mu_1 \geq \mu_2 \quad \dots \geq \mu_{k_1} \geq \mu_{k_1+1} \geq \dots \geq \mu_k$.

If we denote the likelihood ratio by Λ, then we would reject H_0 in favor of H_A for large values of the statistic $L = -2 \log \Lambda$. We first introduce some notation.

We take a geometric approach, and use the symbols H_0 and H_A to represent not only statistical hypotheses but also subspaces of k-dimensional Euclidean space, E^k. The subspaces denoted by H_0 and H_A are closed convex cones. If $\underline{\bar{x}} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k)$ is the vector of observed sample means obtained from taking samples of size n_i: $i=1,2,\ldots k$ from k normal populations with known variances σ_i^2, then the maximum likelihood estimate of $\underline{\mu} = (\mu_1, \mu_2, \ldots \mu_k)$ under $H_A(H_0)$ is the projection of $\bar{x}$ onto the closed convex cone $H_A(H_0)$, where

$$H_0 = \{y \in E^k;\ y_1 = y_2 = \cdots = y_{k_1} \geq y_{k_1+1} = \cdots = y_k\}$$

$$H_A = \{y \in E^k:\ y_1 \geq y_2 \geq \cdots \geq y_k\},$$

the projection being taken with respect to weights $w_i = \dfrac{\sigma_i^2}{n_i}$ (see Barlow et al., Chapter 2). This projection we denote by $E_{\underline{w}}(\underline{\bar{x}}|H_A)$ $(E_{\underline{w}}(\underline{\bar{x}}|H_0))$. The i^{th} component of this vector we denote by $E_{\underline{w}}(\underline{\bar{x}}|H_A)_i$ $(E_{\underline{w}}(\underline{\bar{x}}|H_0)_i)$.

The statistic $L = -2 \log \Lambda$ may be written as

$$(2.3) \qquad L = \sum_{i=1}^{k} [\bar{x}_i - E_{\underline{w}}(\bar{x}|H_0)_i]^2\, w_i - \sum_{i=1}^{k} [\bar{x}_i - E(\underline{\bar{x}}|H_A)_i]^2\, w_i .$$

Using the notation $\|\cdot\|_w$ to denote the norm in E^k associated with the inner product $(x,y)_{\underline{w}} = \sum_{i=1}^{k} x_i y_i w_i$, we may denote L by

$$(2.4) \qquad L = \|\underline{\bar{x}} - E(\underline{\bar{x}}|H_0)\|_{\underline{w}}^2 - \|\underline{\bar{x}} - E(\underline{\bar{x}}|H_A)\|_{\underline{w}}^2 .$$

Using well known properties of projections on closed convex cones (see Barlow et al., Chapter 2) together with some algebra, L may be expressed equivalently as

$$(2.5)\qquad L = \|F\,(\underline{\bar{x}}|H_A)\|^2_{\underline{w}} - \|E_{\underline{w}}(\underline{\bar{x}}|H_0)\|^2_{\underline{w}} .$$

The following inequality is fundamental to the results that follow and we state it as a theorem.

THEOREM 2.1. If $\underline{\delta} \in H_0$, then for any $\underline{x} \in E^k$ we have

$$(2.6)\qquad L(\underline{x} + \underline{\delta}) \geq L(\underline{x}).$$

PROOF. We break up the proof into two separate cases. In the first we consider the case in which

$$(2.7)\qquad \frac{\sum_{i=1}^{k_1} w_i x_i}{\sum_{i=1}^{k_1} w_i} \geq \frac{\sum_{i=k_1+1}^{k} w_i x_i}{\sum_{i=k_1+1}^{k} w_i} ,$$

the weighted average on the left hand side of the inequality sign being the common value of $E_{\underline{w}}(\underline{x}|H_0)_i$, $i = 1,2,\ldots,k$, and the weighted average on the right hand side being the common value of $E_{\underline{w}}(\underline{x}|H_0)_i$, $i = k_1+1,k_1+2,\ldots,k$. We can, without loss of generality assume that $\delta = (\delta,\delta,\ldots\delta,0,0,0,\ldots0)$, where $\delta > 0$, and we have

$$(2.8)\qquad \frac{\sum_{i=1}^{k_1} w_i(x_i + \delta)}{\sum_{i=1}^{k_1} w_i} \geq \frac{\sum_{i=k_1+1}^{k} w_i x_i}{\sum_{i=k_1+1}^{k} w_i}$$

The quantity on the left hand side being the common value of $E_{\underline{w}}(x + \delta | H_0)_i$, $1 \leq i \leq k_1$. It is straightforward to show that

$$(2.9) \qquad \|(\underline{x} + \underline{\delta}) - E_{w}(\underline{x} + \delta | H_0)\|^2_{\underline{w}} = \|\underline{x} - E_{w}(\underline{x} | H_0)\|^2_{\underline{w}} .$$

We now utilize Theorem 2.1 of Robertson and Wegman (1978), which states that if C is any closed convex cone in Hilbert Space, and if $\underline{z} \in C$, then for any $\underline{y}$,

$$(2.10) \qquad \|(\underline{y} + \underline{z}) - E_{\underline{w}}(\underline{y} + \underline{z} | C)\|_{\underline{w}} \leq \|\underline{y} - E_{\underline{w}}(\underline{y} | C)\|_{\underline{w}} .$$

Since $\underline{\delta} \in H_0 \subset H_A$, we have

$$(2.11) \qquad -\|(\underline{x} + \underline{\delta}) - E_{\underline{w}}(\underline{x} + \underline{\delta} | H_A)\|^2 \geq -\|\underline{x} - E_{w}(\underline{x} | H_A)\|^2 .$$

Combining (2.9) and (2.11) we obtain, for the case in question, $L(\underline{x} + \underline{\delta}) \geq L(\underline{x})$.

We now consider the case where x is a vector such that

$$(2.12) \qquad \frac{\sum_{i=1}^{k_1} w_i x_i}{\sum_{i=1}^{k_1} w_i} < \frac{\sum_{i=k_1+1}^{k} w_i x_i}{\sum_{i=k_1+1}^{k} w_i} .$$

In this case the $E_{\underline{w}}(\underline{x} | H_0)_i$, $i=1,2,\ldots,k_1,\ldots,k$, will have the common value $\dfrac{\sum_{i=1}^{k} w_i x_i}{\sum_{i=1}^{k} w_i}$. Let ϵ be the positive difference between the right

and left hand sides of (2.12) and let $\underline{\varepsilon}$, $\underline{\varepsilon} \in H_0$ be the vector with ε as its first k_1 components and 0's elsewhere. We shall show first that if $\underline{\delta}_1 = (\delta_1, \delta_1, \ldots, \delta_1, 0, \ldots, 0)$ and $0 \leq \delta_1 \leq \varepsilon$, then

$$(2.13) \qquad L(\underline{x} + \underline{\delta}_1) \geq L(x).$$

We shall then in fact be done, since $\underline{x} + \underline{\varepsilon}$ is a vector such (2.7) holds, and we will then have for any $\underline{\delta} \in H_0$ that

$$(2.14) \quad L(\underline{x} + \underline{\delta} + \underline{\varepsilon}) \geq L(\underline{x} + \underline{\varepsilon}) \geq L(\underline{x}).$$

Thus all possible cases will be covered.

To prove (2.13) we need two lemmas. The first is obvious, and we state it without proof.

LEMMA 2.2. Let $\{x_1, x_2, \ldots, x_k\}$ and $\{y_1, y_2, \ldots, y_k\}$ be two sets of numbers such that $x_1 \geq x_2 \geq \ldots \geq x_k$, $y_1 \geq y_2 \geq \ldots \geq y_k$ and

$$(2.15) \qquad y_i - y_{i+1} \geq x_i - x_{i+1} \geq 0 .$$

Then for any set of positive weights $\{w_1, w_2, \ldots, w_k\}$

$$(2.16) \qquad \sum_{i=1}^{k} w_i \, (y_i - \bar{y})^2 \geq \sum_{i=1}^{k} w_i \, (x_i - \bar{x})^2$$

where $\bar{x}$ and $\bar{y}$ are averages taken with respect to the weights.

LEMMA 2.3. If $\bar{x}$ is any vector in E^k, and if $\delta \in H_A$, then

$$(2.17) \qquad E_{\underline{w}}(\underline{x} + \delta | H_A)_i - E_{\underline{w}}(\underline{x} + \underline{\delta} | H_A)_{i+1} \geq E_w(\underline{x} | H_A)_i - E_{\underline{w}}(\underline{x} | H_A)_{i+1}$$

$$i = 1, 2, \ldots, k-1.$$

PROOF. The proof follows from noting that the sets on which $E_{\underline{w}}(\underline{x} + \underline{\delta} | H_A)$ assumes constant values are subsets of the sets on which $E_{\underline{w}}(\underline{x} | H_A)$ assumes constant values, and then using the maximum upper sets algorithm. (See Section 2.3 of [Barlow et al., 1972]).

We now return to inequality (2.13), and note that in the particular case we are dealing with, using the form (2.5), $L(\underline{x} + \underline{\delta}_1)$ and $L(\underline{x})$ may be written respectively as the weighted sum of squares of the $E_{\underline{w}}(\underline{x} + \underline{\delta}_1 | H_A)_i$ and the $E_{\underline{w}}(\underline{x} | H_A)_i$ around their weighted means. (We also use the fact, to be found in Barlow et al., that for any $\underline{x} \in E^k$, and any closed convex cone, H, $\sum_{i=1}^{k} w_i x_i = \sum_{i=1}^{k} w_i E_{\underline{w}}(\underline{x} | H)_i$).

Thus we may invoke Lemmas 2.2 and 2.3 to obtain

$$L(\underline{x} + \delta_1) \geq L(\underline{x}),$$

and the proof is completed.

The inequality obtained in Theorem 2.1 enables us to obtain the following result regarding the distribution of the likelihood ratio based statistic, $L(\underline{\bar{X}})$, over the null hypothesis region.

THEOREM 2.4. Define the hypotheses H_0' and H_A' as follows

$$(2.18) \qquad H_0'\colon \mu_1 = \mu_2 = \cdots = \mu_{k_1} \;;\; \mu_{k_1+1} = \cdots = \mu_k$$

$$(2.19) \qquad H_A'\colon \mu_1 \geq \mu_2 \geq \cdots \geq \mu_{k_1} \;;\; \mu_{k_1+1} \geq \cdots \geq \mu_k .$$

Let $L'(\underline{\bar{X}}) = -2 \log \Lambda'(\underline{\bar{X}})$, $\Lambda'(\underline{\bar{X}})$ being the likelihood ratio for testing H_0' against $H_A' - H_0'$. Then

$$\sup_{\mu \in H_0} P_\mu [L > t] = P[L' > t] .$$

PROOF. It is clear from the inequality obtained in Theorem 2.1 that if $\underline{\mu} = (\mu,\mu,\ldots,\mu,0,0,\ldots 0)$, $\underline{\mu} \in H_0$ and $\mu > 0$, then $L(\underline{x} + \underline{\mu})$ is an increasing function of μ, for $\underline{x}$ fixed. Since

$$(2.20) \qquad P_{\underline{\mu}}[L(\underline{\bar{X}}) > t] = P_{\underline{0}}[L(\underline{\bar{X}} + \underline{\mu}) > t] \geq P_{\underline{0}}[L(\underline{\bar{X}}) > t],$$

we note that for all t, $P_\mu[L(\underline{\bar{X}}) > t]$ is also an increasing function of μ. Thus we may state that

$$(2.21) \qquad \sup_{\mu \in H_0} P_\mu[L(\underline{\bar{X}}) > t] = \lim_{\mu\to\infty} P_\mu[L(\underline{\bar{X}}) > t].$$

It is easy to show that the expression on the right is equal to $P[L'(\underline{\bar{X}}) > t]$. To see this we observe that for any given $\underline{x}$, $L(\underline{x})$ and $L'(\underline{x})$ will coincide if the minimum of x_i, $i \leq k_1$ is greater than the maximum of the x_i, $i > k_1$. As μ approaches infinity, the probability of this event approaches 1. Hence we may state

$$(2.22) \qquad \sup_{\mu \in H_0} P_\mu[L(\underline{\bar{X}}) > t] = \lim_{\mu\to\infty} P_\mu[L(\underline{\bar{X}}) > t] = P[L'(\underline{\bar{X}}) > t].$$

The distribution of $L'(\underline{\bar{X}})$ is easy to derive, and is independent of $\underline{\mu} \in H_0$. To see this we let $L_1(\underline{\bar{X}})$ and $L_2(\underline{\bar{X}})$ be the likelihood ratio based statistics for testing

$$H_{01}: \mu_1 = \ldots = \mu_{k_1}$$

against

$$H_{A1}: \mu_1 \geq \mu_2 \geq \cdots \geq \mu_{k_1}$$

and

$$H_{02}: \mu_{k_1+1} = \cdots = \mu_k$$

against

$$H_{A2}: \mu_{k_1+1} \geq \cdots \geq \mu_k$$

($L_1(\underline{\bar{X}}) = -2 \log \Lambda_1(\underline{\bar{X}})$, and $L_2(\underline{\bar{X}}) = -2 \log \Lambda_2(\underline{\bar{X}})$, the Λ_i being the respective likelihood ratios). Clearly $L'(\underline{\bar{X}}) = L_1(\underline{\bar{X}}) + L_2(\underline{\bar{X}})$. L_1 and L_2 are independent, since L_1 is a function of the first k_1 variables only, and L_2 a function of the second $k-k_1$. Furthermore, both L_1 and L_2 are distributed according to Bartholomew's chi-bar-squared distribution (see Barlow et at., Chapter 2).

Arguing as in Section 5 of Robertson and Wegman (1978) we find that the null hypothesis distribution of L' is a chi-bar-square distribution where the coefficients on the various chi-square tail probabilities can be found by convoluting two sequences of level probabilities corresponding to linear order restrictions. Specifically

(2.23) $P[L' = 0] = P_1(1,k_1)\ P_2(1,k-k_1)$

and

(2.24) $P[L' > t] = \sum_{i=3}^{k} Q_i\ P[\chi^2_{i-2} > t]\ ;\ t > 0$

where $P_1(\ell_1,k_1)$ is the probability that $E_w(\underline{\bar{X}}|H_{A1})$ assumes ℓ_1 distinct values amongst its first k_1 components and $P_2(\ell_2,k-k_1)$ is the probability that $E_w(\underline{\bar{X}}|H_{A2})$ assumes ℓ_2 distinct values amongst

its second $k-k_1$ components and the sequence $\{Q_i\}_{i=2}^{k}$ is the convolution of the two sequences $\{P_1(\ell_1,k_1)\ ;\ \ell_1 = 1,2,\ldots,k_1\}$ and $\{P_2(\ell_2,k-k_1);\ \ell_2 = 1,2,\ldots,k-k_1\}$. The probabilities $P_1(\ell_1,k_1)$ and $P_2(\ell_2,k-k_1)$ are computed und the assumption that H_{01} and H_{02} are true. Algorithms for their computation are discussed at length in [Barlow et al., 1972] and an approximation is given in [Robertson and Wright, 1983].

Thus far we have dealt only with the case in which the null hypothesis postulates a descending trend between two blocks of parameters. We may in fact extend our result to cover an arbitrary number of blocks of parameters. For example consider the problem of testing H_0 against H_A-H_0 where

(2.25) H_0: $\mu_1 = \mu_2 = \cdots = \mu_{k_1} \geq \mu_{k_1+1} = \cdots = \mu_{k_2} \geq \cdots \geq \mu_{k_{\ell-1}+1} = \cdots = \mu_k$

(2.26) H_A: $\mu_1 \geq \cdots \geq \mu_{k_1} \geq \mu_{k_1+1} \geq \cdots \geq \cdots \geq \mu_k$.

Here the null hypothesis postulates equality within ℓ sets of parameters, and a descending trend between blocks, while H_A postulates a descending trend both within and between blocks.

Theorems 2.1 and 2.4 have natural extensions for this situation, and in that regard we define the hypotheses H_0' and H_A' as follows:

(2.27) H_0': $\mu_1 = \mu_2 = \cdots = \mu_{k_1};\ \mu_{k_1+1} = \cdots = \mu_{k_2}\ ;\ldots;\ldots;\ \mu_{k_{\ell-1}+1} = \cdots = \mu_k$

(2.28) H_A': $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_{k_1};\ \mu_{k_1+1} \geq \cdots \geq \mu_{k_2}\ ;\ldots;\ldots;\ \mu_{k_{\ell-1}+1} \geq \cdots \geq \mu_k$

Analogous to the inequality for two blocks proved in Theorem 2.1, we may use induction to obtain for each $\underline{x}$,

$$(2.29) \quad L(\underline{x}) \le L'(\underline{x})$$

where L and L' are the likelihood ratio based statisticis for testing H_0 against $H_A - H_0$ and H_0' against $H_A' - H_0'$, respectively.

Similarly, using the inequality we may show that

$$(2.30) \quad \sup_{\underline{\mu} \in H_0} P_{\underline{\mu}}\{L(\underline{X}) > t\} = P\{L'(\underline{X}) > t\}.$$

The distribution of $L'(\bar{\underline{X}})$ may be easily shown to be distributed as the sum of ℓ independent random variables each distributed according to a version of Barthomew's chi-bar-square distribution. Thus, we may state

$$(2.31) \quad P[L'(\bar{\underline{X}}) > t] = \sum_{j=\ell+1}^{k} Q_j \, P[\chi^2_{j-1} > t] \qquad \text{For} \quad t > 0$$

$$(2.32) \quad P[L'(\bar{\underline{X}}) = 0] = \prod_{i=1}^{\ell} P_i(1, k_i - k_{i-1})$$

where the sequence $\{Q_j\}_{j=\ell}^{k}$ is found by convoluting ℓ sequences of level probabilities as in (2.23) and (2.24).

3. Asymptotic tests for the exponential family and the multinomial distribution.

Because the analysis in this section so closely parallels the work in [Robertson and Wegman, 1978] and in [Robertson, 1978] we present only sketches of the arguments. We first consider an exponential family $\{f(\cdot;\theta,\tau)\}$ which, following Lehmann, 1959 , we parameterize as follows.

(3.1) $f(x;\ \theta,\tau) = \exp[p_1(\theta)p_2(\tau)k(x) + s(x,\tau) + q(\theta;\tau)];\ \theta \in (\theta_1,\theta_2);\ \tau \in T$

The following assumptions are made:

(3.2) $p_1(\cdot)$ and $q(\cdot;\tau)$ both have continuous second derivatives on some interval (θ_1,θ_2) for all $\tau \in T$.

(3.3) $p_1'(\theta) > 0$ for all $\theta \in (\theta_1,\theta_2)$ and for all $\tau \in T$.

(3.4) $q'(\theta;\tau) = -\theta\, p_1'(\theta)p_2(\tau)$ for all $\theta \in (\theta_1,\theta_2)$ and for all $\tau \in T$.

We are thinking of τ as fixed so that all derivatives are with respect to θ. Using this parameterization we have

(3.5) $E[k(X)] = \theta$

(3.6) $\mathrm{var}\,[k(X)] = [p_1'(\theta)p_2(\tau)]^{-1}$.

We consider the problem of testing H_0 against H_A-H_0 where we sample from each of k independent population and

$$H_0:\ \theta_1 = \theta_2 = \dots = \theta_{k_1} \geq \theta_{k_1+1} = \dots = \theta_k$$

$$H_A:\ \theta_1 \geq \theta_2 \geq \dots \geq \theta_{k_1} \geq \theta_{k_1+1} \geq \dots \geq \dots \geq \theta_k$$

(In this section we consider only two blocks. Extensions of the results presented here to an arbitrary number of blocks are straightforward. In the analysis of the unipolar affective disorder data in Section 4 the distribution is binomial.)

Suppose we take a sample of size n_i from the i^{th} population: $i = 1,2,\ldots,k$, and letting Λ be the likelihood ratio, we have

$$(3.7)\qquad L = -2 \log \Lambda$$

$$= 2 \sum_{i=1}^{k} \{n_i \hat{\theta}_i p_2(\tau_i)[p_1(\hat{\theta}_i) - p_1(E_{\underline{w}}(\hat{\theta}|H_0)_i)]$$

$$+ n_i[q(\hat{\theta}_i;\tau_i) - q(E_{\underline{w}}(\hat{\theta}|H_0)_i;\ \tau_i)]\}$$

$$-2 \sum_{i=1}^{k} \{n_i \hat{\theta}_i p_2(\tau_i)[p_1(\hat{\theta}_i) - p_1(E_{\underline{w}}(\hat{\underline{\theta}}|H_A)_i]$$

$$+ n_i[q(\hat{\theta}_i;\tau_i) - q(E_{\underline{w}}(\hat{\underline{\theta}}|H_A)_i;\ \tau_i)]\}$$

where

$$(3.8)\qquad \hat{\theta}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} k(X_{ij})$$

and the weights, w_i, are given by

$$(3.9)\qquad w_i = [n_i p_2(\tau_i)] \cdot [\sum_{j=1}^{k} n_j p_2(\tau_j)]^{-1}.$$

Using a second degree Taylor series expansion, and equation (3.4) we may write L in the form

$$(3.10)\qquad L = -\{\sum_{i=1}^{k} n_i[\hat{\theta}_i p_2(\tau_i) p_1''(a_i) + q''(b_i;\tau_i)] \cdot [\hat{\theta}_i - E_{\underline{w}}(\hat{\underline{\theta}}|H_0)_i]^2\}$$

$$- \{-\sum_{i=1}^{k} n_i[\hat{\theta}_i p_2(\tau_i) p_1''(c_i) + q''(d_i;\tau_i)] \cdot [\hat{\theta}_i - E_{\underline{w}}(\hat{\underline{\theta}}|H_A)_i]^2\}$$

Where a_i and b_i lie between $\hat{\theta}_i$ and $E_{\underline{w}}(\hat{\theta}|H_0)_i$, and c_i and d_i lie between $\hat{\theta}_i$ and $E_{\underline{w}}(\hat{\theta}|H_A)_i$.

Using the results of the previous section we obtain the following result regarding the asymptotic distribution of L.

Theorem 3.1

$$(3.11) \quad \sup_{\underline{\theta} \in H_0} \lim_{n_i \to \infty} P[L > t] = P[T' > t]$$

where

$$T' = \sum_{i=1}^{k} \gamma_i p_1'(\theta_i) p_2(\tau_i)[X_i - E_{\underline{w}}(\underline{X}|H_0')_i]^2$$

$$- \sum \gamma_i p_1' \theta_i p_2(\tau_i)[X_i - E_w(\underline{X}|H_A')_i]^2$$

and where $\underline{X} = (X_1, X_2, \ldots, X_k)$ is a k dimensional vector of normally distributed independent random variables with

$$(3.12) \quad E(X_i) = 0$$

and

$$(3.13) \quad \mathrm{var}\,(X_i) = [\gamma_i p_i'(\theta_i) p_2(\tau_i)]^{-1} \quad .$$

H_0' and H_A' are the closed convex cones associated with the hypotheses

$$H_0': \ \theta_1 = \theta_2 = \dots = \theta_{k_1};\ \theta_{k_1+1} = \dots = \theta_k$$

$$H_A': \ \theta_1 \geq \theta_2 \geq \dots \geq \theta_{k_1};\ \theta_{k_1+1} \geq \dots \geq \theta_k.$$

The constants γ_i are defined in the following way. We assume we are dealing with sets of sample sizes, $\{n_i\}$, $i=1,\dots,k$, such that for each set, $\frac{n_i}{n_j}$ is some constant, c_{ij}. We define

$$\gamma_i = \frac{n_i}{\min\limits_{1\leq i\leq k} \{n_i\}} \quad .$$

PROOF. We first consider some arbitrary $\underline{\theta} \in H_0$, and we define hypotheses H_0'' and H_A'' by dividing the hypotheses H_0 and H_A in separate parts according to the different values assumed by $\underline{\theta}$. This is most easily explained by an example: suppose we are testing

$$H_0: \ \theta_1 = \theta_2 \geq \theta_3 = \theta_4$$

against

$$H_A: \ \theta_1 \geq \theta_2 \geq \theta_3 \geq \theta_4$$

and suppose that $\underline{\theta}$ is such that

$$\theta_1 = \theta_2 > \theta_3 = \theta_4,$$

then we define H_0'' and H_A'' as follows

$$H_0'' : \ \theta_1 = \theta_2 \ ; \ \theta_3 = \theta_4$$

$$H_A'' : \ \theta_1 \geq \theta_2 \ ; \ \theta_3 \geq \theta_4$$

If $\theta_2 = \theta_3$ then $H_0'' = H_0$ and $H_A'' = H_A$.

Now, using techniques similar to those in [Robertson and Wegman, 1978] it follows that

$$L \Longrightarrow \sum_{i=1}^{k} \gamma_i P_i(\theta_i) P_2(\tau_i) [\ \underline{X}_i - E_{\underline{w}}(\underline{X}|H_0'')_i]^2$$

$$- \sum_{i=1}^{k} \gamma_i P_i(\theta_i) P_2(\tau_2) [\ \underline{X}_i - E_{\underline{w}}(\underline{X}|H_A'')_i]^2$$

where $\underline{X}$ and $\underline{w}$ are as described in the theorem. We denote this limiting random variable by T''. Using the results obtained in this paper for normal random variables we conclude that

$$\lim_{n \to \infty} P_{\underline{\theta}}[L' > t] = P_{\underline{\theta}}[T'' > t] \leq P[T' > t].$$

To show that the supremum is attained, note that if $\theta_{k_1} > \theta_{k_1+1}$ then H_0'' coincides with H_0', H_A'' coincides with H_A', and $T'' = T'$.

In Section 4 we use Theorem 3.1 to treat the unipolar affective disorder data mentioned in the introduction.

We now treat the case in which we are sampling from a multinomial distribution, i.e. we have a random vector $\underline{X}$ where

$$f_{\underline{X}}(x_1,\ldots,x_k) = \frac{n!}{x_1!\ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}, \quad \sum_{i=1}^{k} x_i = n, \quad \sum_{i=1}^{k} p_i = 1 .$$

If we wish to maximize the likelihood subject to the restriction that the p_i satisfy the partial order, the maximizing estimates may be obtained by taking the unrestricted maximum likelihood estimates, $\hat{p}_i = \frac{x_i}{n}$, and projecting them onto the closed convex cone associated with the partial order. The projection is taken using unit weights (see Example 2.1 in [Barlow et al., 1972]).

We treat the problem of testing H_0 against $H_A - H_0$ where

$$H_0:\ p_1 = p_2 = \cdots = p_{k_1} \geq p_{k_1+1} = \cdots = p_k$$

$$H_A:\ p_1 \geq p_2 \geq \cdots \quad \cdots \geq p_k.$$

The treatment here closely parallels the work in Robertson (1978).

If Λ is the likelihood ratio , and we let $L = -2 \log \Lambda$, then we have

$$(3.14) \quad L = 2\sum_{i=1}^{k} n\hat{p}_i[\ln E(\underline{\hat{p}}|H_A)_i] - 2\sum_{i=1}^{k} n\hat{p}_i[\ln E(\underline{\hat{p}}|H_0)_i]$$

where $E(\underline{\hat{p}}|H_A)$ and $E(\underline{\hat{p}}|H_0)$ denote the projections, using equal weights, of $\underline{\hat{p}}$ onto H_A and H_0 respectively. Using second degree Taylor series expansion we may rewrite this as

$$(3.15)\quad L = \sum_{i=1}^{k} n\hat{p}_i \frac{1}{b_i^2}[\hat{p}_i - E(\hat{p}|H_0)_i]^2 - \sum_{i=1}^{k} n\hat{p}_i \frac{1}{a_i^2}[\hat{p}_i - E(\hat{p}|H_A)_i]^2$$

where b_i lies between $\hat{p}_i$ and $E(\underline{\hat{p}}|H_0)_i$, and a_i lies between $\hat{p}_i$ and $E(\underline{\hat{p}}|H_A)_i$. Writing L in this form enables us to prove the following theorem, analagous to Theorem 3.1 for the exponential family:

THEOREM 3.2

$$(3.20)\quad \sup_{\underline{p}\in H_0} \lim_{n\to\infty} P_{\underline{p}}[L > t] = P[T > t]$$

where

$$(3.21)\quad T = \sum_{i=1}^{k} p_i[Z_i - E_p(\underline{Z}|H_0')_i]^2 - \sum_{i=1}^{k} p_i[Z_i - E_{\underline{p}}(\underline{Z}|H_A')_i]^2 ,$$

$\underline{Z} = (Z_1,\ldots Z_k)$ being a vector of k independent normally distributed random variables with

$$(3.22)\quad E(Z_i) = 0$$

$$(3.23)\quad \mathrm{var}\,(Z_i) = p_i^{-1} .$$

The hypotheses H_0' and H_A' are defined by the closed convex cones associated with

$$H_0':\ p_1 = p_2 = \ldots = p_{k_1};\ p_{k_1+1} = \ldots = p_k$$

$$H_A':\ p_1 \geq p_2 \geq \ldots \geq p_{k_1};\ p_{k_1+1} \geq \ldots \geq p_k$$

PROOF. The proof is similar to the proof of Theorem 3.1 and closely parallels the proof of Theorem 2 in [Robertson, 1978].

4. An example involving the binomial distribution.

We now consider the problem discussed in Section 1 concerning the two theories which explain the age of onset of unipolar affective disorder. The hypotheses in question may be stated as

$$H_0: \; p_1 = p_2 = p_3 \geq p_4 = p_5 = p_6$$

$$H_A: \; p_1 \geq p_2 \geq p_3 \geq p_4 \geq p_5 \geq p_6$$

where H_0 represents the "qualitative" theory of genetic transmission

and H_A the "quantitative" theory. From a mathematical point of view, the most interesting data set in Cadoret et al. regards the proportion of depressive parents in males suffering from u.a.d.. For these data the six sample sizes (i.e. numbers of male probands in each group) are 66, 174, 134, 166, 116, and 140 and the six relative frequencies are .212, .103, .149, .145, .086, .100. Any one of several algorithms in Barlow et al., 1972 can be used to compute the maximum likelihood estimates restricted by H_0 and H_A. Using the pool adjacent violators algorithm, we found that the maximum likelihood estimates subject to H_0 are .139, .139, .139, .114, .114, .114 and the estimates satisfying H_A are .212, .131, .131, .131, .131, .094, .094. The resulting likelihood ratio has a value of 5.362.

In order to compute a P-valve we must compute

$$\sum_{3 \leq \ell_1+\ell_2 \leq 6} P_1(\ell_1,3)P_2(\ell_2,3)\ P\,[\chi^2_{\ell_1+\ell_2-2} \geq 5.362].$$

The values of the level probabilities, $P_1(\ell_1 3)$ and $P_2(\ell_2,3)$ depend upon the sample sizes (weights) and can be found using the formulas on page 140 of Barlow et. al (1972). The resulting P-value is .0648.

Additional details regarding the analysis of this data set and analysis of the other data from Cadoret et. al may be found in Robertson and Warrack (1981).

5. Comments.

As noted in the introduction, the null hypothesis behavior of the distribution of the likelihood ratio test statistic is not what one might expect in light of previous research in the area of testing

statistical hypotheses under order restriction. Moreover, the test statistic L', which is a likelihood ratio statistic for testing H_0': $p_1 = p_2 = p_3$, $p_4 = p_5 = p_6$ against $H_A' - H_0'$ where H_A': $p_1 \geq p_2 \geq p_3$, $p_4 \geq p_5 \geq p_6$, has H_0'(and thus H_o) as a similar region and has exactly the same null hypothesis distribution as L. This, together with the fact that $L \leq L'$ means that L' is uniformly more powerful than L. This result seems somewhat paradoxical in that L' does not account for the prior knowledge that $p_3 \geq p_4$. If we use L' to compute a P-valve for our data set then the observed valve is 5.783 yielding a P-valve of .0533.

In computing P-values, power considerations may not be too pertinent since P-values are computed under the null hypothesis and power says something about the quality of the test when the alternative is true. It seems to us that the P-value computed using L is rather more representative of the evidence in the data against H_0 than the P-value computed using L'.

References.

1). Barlow, R.E., Bartholomew, D.J., Bremner, J.M., and Brunk, H.D. (1972) Statistical Inference Under Order Restrictions. Wiley, New York.

2). Cadoret, Remi J., Woolson, Robert, and Winokur, George (1977), Relationship of age of onset in unipolar affective disorder to risk of alcoholism and depression in parents, Journal of Psychiatric Research, 13, 137-142.

3). Dykstra, R.L. and Robertson, Tim (1982), Order restricted statistical tests on multinomial and Poisson parameters: the starshaped restriction, Annals of Statistics, 1246-1252.

4). Dykstra, R.L. and Robertson, Tim (1983), On testing monotone tendencies, Journal of the American Statistical Association, 78, (to appear).

5). Robertson, Tim (1975), On likelihood ratio tests of trend, University of North Carolina Institute of Statistics Mimeo Series No. 991.

6). Robertson, Tim (1978), Testing for and against an order restriction on multinomial parameters, Journal of the American Statistical Association, 73, 197-202.

7). Robertson, Tim and Warrack, Giles (1981), On testing two theories regarding the genetic makeup of patients suffering from unipolar affective disorder, University of Iowa, Department of Statistics and Actuarial Science Technical Report No. 76.

8). Robertson, Tim and Wegman, Edward J. (1978), Likelihood ratio tests for order restriction in exponential families, Ann. Statist., 6, 485-505.

9). Robertson, Tim and Wright, F.T. (1981), Likelihood ratio tests for and against a stochastic ordering between multinomial populations, Annals of Statistics, 9, 1248-1257.

10). Robertson, Tim and Wright, F.T. (1983), On approximation of the level probabilities and associated distributions in order restricted inference, Biometrika, 70, (to appear).

11). Winokur, G., Cadoret, R.J., Dorzab, J. and Baber, M. (1971), Depressive disease. A genetic study, Archsgen Psychist, 24, 135.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 13	2. GOVT ACCESSION NO. AD-A123 746	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Likelihood Ratio Test Regarding Two Nested But Oblique Order Restricted Hypotheses		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Giles Warrack and Tim Robertson		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics & Actuarial Science The University of Iowa		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Department of the Navy Arlington, VA		12. REPORT DATE November 1982
		13. NUMBER OF PAGES 25
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Order restricted inference, likelihood ratio tests, unipolar affective disorder, chi-bar-square distribution, isotonic regression.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

>In an article in the Journal of Psychiatric Research, Cadoret, Woolson, and Winokur (1977) consider two theories regarding the genetic makeup of patients suffering from unipolar affective disorder. These two theories imply nested but oblique order restrictions on the parameters of a statistical model. A likelihood ratio test for these two restrictions is studied.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

LMED

-8